

# significance

quantitative research



Demis

# Strategisch Personenauto Rekenkader (SPARK)

## Kwaliteitsrapport

3e versie | mei 2023

# Strategisch Personenauto Rekenkader (SPARK)

## Kwaliteitsrapport

3e versie | 31 mei 2023

**Auteurs:**

Gerard de Jong, Marco Kouwenhoven, Michiel Benjamins

**Projectnummer:**

20025

### Toegankelijkheid

Het PBL hecht veel waarde aan de toegankelijkheid van zijn producten. Mocht u problemen ervaren bij het lezen van dit document, dan kunt u contact opnemen via [info@pbl.nl](mailto:info@pbl.nl). Vermeld daarbij s.v.p. de naam van deze publicatie ("Kwaliteitsrapport SPARK"), het paginanummer en het probleem waar u tegenaan loopt.

# Inhoudsopgave

<b>1. Inleiding</b>	<b>5</b>
<b>2. Kwaliteit van de modelschattingen</b>	<b>7</b>
2.1 Waar hangt de kwaliteit van de modelschattingen van af?	7
2.2 Criteria voor de kwaliteit van de modelschattingen	7
2.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelschatting	7
2.4 Uitkomsten voor de kwaliteit van de modelschatting	10
<b>3. Kwaliteit van de modelimplementatie</b>	<b>12</b>
3.1 Waar hangt de kwaliteit van de modelimplementatie van af?	12
3.2 Criteria voor de kwaliteit van de modelimplementatie	12
3.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelimplementatie	13
3.4 Uitkomsten voor de kwaliteit van de modelimplementatie	14
<b>4. Kwaliteit van de modelvalidatie</b>	<b>16</b>
4.1 Waar hangt de kwaliteit van de modelvalidatie van af?	16
4.2 Criteria voor de kwaliteit van de modelvalidatie	16
4.3 Uitgevoerde maatregelen en uitkomsten voor de kwaliteit van de modelvalidatie	17
<b>5. Kwaliteit van de modelrapportage</b>	<b>19</b>
5.1 Waar hangt de kwaliteit van de modelrapportage van af?	19
5.2 Criteria voor de kwaliteit van de modelrapportage	19
5.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelrapportage	19
<b>6. Conclusies</b>	<b>21</b>

# 1. Inleiding

Dit is het kwaliteitsrapport van het onderzoek ‘Landelijk personenautoparkmodel’ dat Significance en Demis uitvoeren voor RWS-WVL en PBL. In de loop van dit project heeft het model de naam SPARK gekregen: ‘Strategisch personenauto rekenkader’.

De ontwikkeling van een nieuw autoparkmodel is een omvangrijk project waarbij hoge eisen aan de kwaliteit worden gesteld. Een hoge kwaliteit vermindert direct ook de risico’s op niet plausibele uitkomsten van het model. Om dit kwaliteitsniveau te bereiken zijn in het project diverse maatregelen genomen. Hierbij gaat het zowel om de inhoudelijke kwaliteit (schattingen, werking van het model, software, rapporten) en de procesmatige kwaliteit.

Naast een projectteam bij Significance en Demis dat veel ervaring heeft bij de ontwikkeling van grote modelsystemen vergelijkbaar met dit nationale autoparkmodel en specifiek ook met autobezits- en autogebruiksmodellen, is gebruik gemaakt van diverse experts van buiten het projectteam. Hierbij ligt er de nadruk op dat voor de inhoudelijke kwaliteit het belangrijk is dat de opzet en de uitwerking van het nieuwe modelsysteem getoetst worden met andere experts.

In de eerste plaats is dit gebeurd door afstemming met de directe begeleiders namens de opdrachtgevers WVL en PBL (waaronder wekelijks voortgangsoverleg, maar ook beoordeling van de diverse rapporten in dit project):

- Dr. Konstanze Winter – projectleider namens WVL;
- Ir. Remko Smit – WVL;
- Jordy van Meerkerk, MSc. – projectleider namens PBL;
- Dieuwert Blomjous, MSc – PBL.

Daarnaast heeft de opdrachtnemer een expertteam aan het eigen projectteam toegevoegd. Op een vijftal momenten in het project (aan het begin van de ontwikkeling van het conceptueel model, halverwege de exploratiefase, aan het begin en op het eind van de schattingsfase en in de loop van de proeftuinfase) is een brainstormsessie gehouden met deze experts. Deze momenten zijn zo gekozen dat de input van deze experts echt gebruikt kon worden in de ontwikkeling van het model en impact had op de beslissingen die hierbij gemaakt worden. Het gaat hier om de volgende inhoudelijke experts namens de opdrachtnemer:

- Prof. Dr. Bert van Wee (TU Delft);
- Dr. Sander van Cranenburgh (TU Delft);
- Dr. Anders Fjendbo Jensen (Danish Technical University, DTU).

De experts zijn met name betrokken bij het formuleren van het conceptuele model en de beleidscontext en -uitkomsten (van Wee), de schatting van keuzemodellen (van Cranenburgh) en de simultane schatting van discrete keuzemodellen en diffusiemodellen (Jensen).

Daarnaast was er ook nog het expertteam van de opdrachtgever dat de kwaliteit van het project heeft beoordeeld. Deze experts gaven een onafhankelijk oordeel over het project als geheel en belangrijke delen afzonderlijk op basis van producten die al in een gevorderd stadium waren. Zij hebben zowel schriftelijk als mondeling commentaar geleverd, wat vervolgens is doorgesproken tijdens een drietal overleggen met de beide experts, de opdrachtgever en de opdrachtnemer. Deze experts namens de opdrachtgever waren:

- Prof. Dr. Henk Meurs (Radboud Universiteit Nijmegen);
- Prof. Dr. Harry Timmermans (TU Eindhoven).

Verder hebben ook andere inhoudelijke experts van WVL, PBL en het Kennisinstituut voor Mobiliteitsbeleid KiM aan de expertbijeenkomsten deelgenomen, en ook diverse projectdocumenten beoordeeld.

Naast de vooral wetenschappelijke toetsing door deze technische experts zijn er ook bijeenkomsten met de beleidsafdelingen op het gebied van verkeersbeleid en fiscaal beleid op de automarkt geweest met vertegenwoordigers van het Ministerie van Infrastructuur en Waterstaat en het Ministerie van Financiën. Hierbij lag de nadruk op twee aspecten van het model:

- De beleidsknoppen aan de invoerkant;
- De uitvoervariabelen van het model.

Verder zijn er gesprekken geweest van de opdrachtgever en de opdrachtnemer met diverse organisaties die zich bezighouden met de automarkt in de praktijk of onderzoek daar naar. Deze gesprekken gingen zowel over de inhoudelijke kennis van de automarkt van deze partijen, mogelijk te gebruiken databronnen als de aspecten die ook met de beleidsafdelingen zijn besproken. De organisaties waarmee in dit kader is overlegd (soms meerdere keren) zijn:

- ANWB;
- CBS;
- RAI/BOVAG;
- RDC;
- RDW;
- Revnext;
- Rijksdienst voor Ondernemend Nederland RVO;
- TNO;
- Vereniging van Nederlandse Autoleasebedrijven VNA;
- Vereniging voor Zakelijke Rijders VZR.

Significance en Demis willen alle genoemde organisaties en personen hartelijk danken voor hun inbreng, die heeft bijgedragen tot de huidige versie van SPARK. Desalniettemin ligt de verantwoordelijkheid voor het ontwikkelde prognosemodel SPARK bij de opdrachtnemer.

Dit rapport beschrijft welke activiteiten in de loop van dit project zijn uitgevoerd om te komen tot een prognosemodel van hoge kwaliteit, op basis van welke criteria de kwaliteit wordt beoordeeld en wat de uitkomsten zijn van deze beoordeling. Hierbij wordt een onderscheid gemaakt naar:

- de kwaliteit van de modelschattingen;
- de kwaliteit van de modelimplementatie;
- de kwaliteit van de modelvalidatie;
- de kwaliteit van de modelrapportages.

Deze vier onderwerpen komen aan de orde in de hoofdstukken 2 tot en met 5. Hierbij wordt steeds besproken waar de kwaliteit van afhangt, volgens welke criteria kwaliteit wordt vastgelegd, wat er in het project gedaan is om de kwaliteit op een hoog niveau te brengen en wat op basis van de criteria voor ieder van de vier onderdelen de vastgestelde kwaliteit is.

Hoofdstuk 6 bevat de meer algemene conclusies over de kwaliteit van het model SPARK.

## 2. Kwaliteit van de modelschattingen

### 2.1 Waar hangt de kwaliteit van de modelschattingen van af?

In dit hoofdstuk trekken we de kwaliteitsbeoordeling ruimer dan de pure modelschatting zelf. We betrekken ook de modelstructuur en modelspecificatie erin en de kwaliteit van de data. Hierbij gaat het dan ook niet alleen om de statistische c.q. wetenschappelijke kwaliteit van de schattingsresultaten, maar ook of het model ‘fit for purpose’ is: of het de belangrijkste exogene ontwikkelingen beschrijft, de beleidsknoppen bevat die nodig zijn voor het doorrekenen van verkeers- en fiscaal beleid inzake de personenautomarkt en de uitvoervariabelen die daarbij gevraagd worden.

De kwaliteit van de modelschatting neemt doorgaans toe als de data van betere kwaliteit zijn (o.a. hogere mate van representativiteit, minder antwoordfouten en codeerfouten) en als er meer waarnemingen zijn (in dit geval geldt ‘meer is beter’). Dit verhoogt de betrouwbaarheid en nauwkeurigheid van de schattingen. Verder is de kwaliteit van de modelschattingen afhankelijk van de keuze voor een goede modelspecificatie, o.a. welke variabelen in het model opgenomen, in welke vorm (b.v. lineair, logaritmisch of in klassen) en of er rekening gehouden wordt met vertekeningen (bias) die in de data zou kunnen zitten.

### 2.2 Criteria voor de kwaliteit van de modelschattingen

Verscheidende modelspecificaties zijn in de schattingsfase beoordeeld en met elkaar vergeleken op de volgende criteria:

- **Modelfit:** voor regressiemodellen gaat het hier om de R-kwadraat; voor discrete keuze modellen en andere modellen die geschat worden met de maximum likelihood methode gaat het om de waarde van de loglikelihoodfunctie en de pseudo-rho-kwadraat. Dit laatste is de waarde van de loglikelihood minimaliseringsfunctie van het geschatte model t.o.v. geen model en t.o.v. een model met alleen constanten voor de alternatieven. Voor vergelijking van modelspecificaties die via lineaire restricties uit elkaar volgen, dient ook de likelihood ratio toets, een formele statistische test die met behulp van de Chi-kwadraat tabel kijkt of het ene model de data significant beter fit dan het andere;
- Komt het teken en de orde van grootte van de geschatte coëfficiënten overeen met de theorie en met a priori verwachtingen? Bij geneste logit modellen gelden er specifieke restricties voor de waarden van de nestcoëfficiënten;
- Precisie/significantie van de schattingen: t-waarden van de coëfficiënten;
- Elasticiteiten voor veranderingen in belangrijke variabelen (met name kosten en inkomens) en overeenstemming hiervan met de range uit literatuur (zoals vastgelegd in het toetsingskader in Memo 02 in dit project). Hier wordt in de validatie (proeftuin) naar gekeken.

### 2.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelschatting

#### Keuze voor een dynamisch model voor iedere tijdshorizon

Allereerst kijken we hierbij of het model goed afgestemd is op de doelstellingen van de opdrachtgevers met het model.

Het doel van dit project is om een prognosemodel te ontwikkelen dat:

- jaar-op-jaar prognoses kan maken voor de korte termijn (1, 2, 3, ... 10 jaar), waarbij veel aandacht moet zijn voor het doorrekenen van de effecten van diverse beleidsmaatregelen;
- langetermijnprognoses kan maken voor over 10, 20, 30 en 40 jaar (en ook tussenliggende jaren).

Jaar-op-jaar prognoses op korte termijn kunnen eigenlijk alleen maar door dynamische modellen goed gemaakt worden. Statische modellen (zowel geaggregeerd als gedesaggregeerd) kunnen dit in het



algemeen heel slecht. Prognoses in de tijd voor verschijnselen die niet zo snel veranderen, zoals het autopark, worden accurater en stabiel door uit te gaan van de bekende huidige verdeling, en uitsluitend de veranderingen daarop te voorspellen. Statische prognosemodellen, die ieder jaar weer ‘from scratch’ beginnen, leveren vaak onrealistisch grote veranderingen in de tijd. Hierdoor zijn statische modellen ook niet goed geschikt voor het doorrekenen van de kortetermijneffecten van beleidsmaatregelen. Dit betreft bijvoorbeeld het effect van veranderingen in diverse componenten van de autokosten op de veranderingen in autobezit in de tijd en specifiek hierop gerichte instrumenten zoals subsidies op vervroegde sloop van auto’s.

De experts namens de opdrachtgever waren het met dit argument eens, en stelden voor om ook de lange termijnprognoses met hetzelfde dynamische model te doen. Hierdoor ontstaat er geen overgang of zelfs sprong ergens op de middellange termijn tussen het dynamische en een lange termijn statisch model. Nadat positieve ervaringen met dit voorstel waren opgedaan in de aanvankelijke schattingen van de dynamische modellen en in de toepassing van het prototype is besloten om op deze weg (dus alleen een dynamisch model) door te gaan. Deze toepassing en toetsing van modelconcepten eerst in een prototype is ook een maatregel die heeft bijgedragen tot de modelkwaliteit.

## Modelstructuur

Met dynamisch model bedoelen we dat voor het aantal auto’s in een huishouden in SPARK de veranderingen van jaar op jaar worden gemodelleerd (b.v. van 1 naar 2 auto’s, maar ook vervanging van een auto). Dit noemen we de ‘transactiemodellen’. Conditioneel op een transactie waarbij een extra auto wordt aangeschaft of een auto vervangen, is er binnen SPARK de typekeuze voor de extra of vervangende auto. Daarnaast wordt het autogebruik gemodelleerd (aantal autokilometers per auto per jaar). Deze drie hoofdkeuzes (bezit, gebruik en type) worden ook onderscheiden in Figuur 1. Verschillende concepten voor de modelstructuur zijn mede op basis van de bestaande literatuur en vanuit de gestelde functionele eisen opgesteld, vastgelegd in notities en rapporten en besproken met de opdrachtgever en de experts namens de opdrachtnemer en de opdrachtgever. Uiteindelijk is op basis van deze gesprekken gekozen voor de modelstructuur zoals afgebeeld in Figuur 1. Binnen de hoofdkeuzes bezit, gebruik en type zijn er aparte modelstructuren voor:

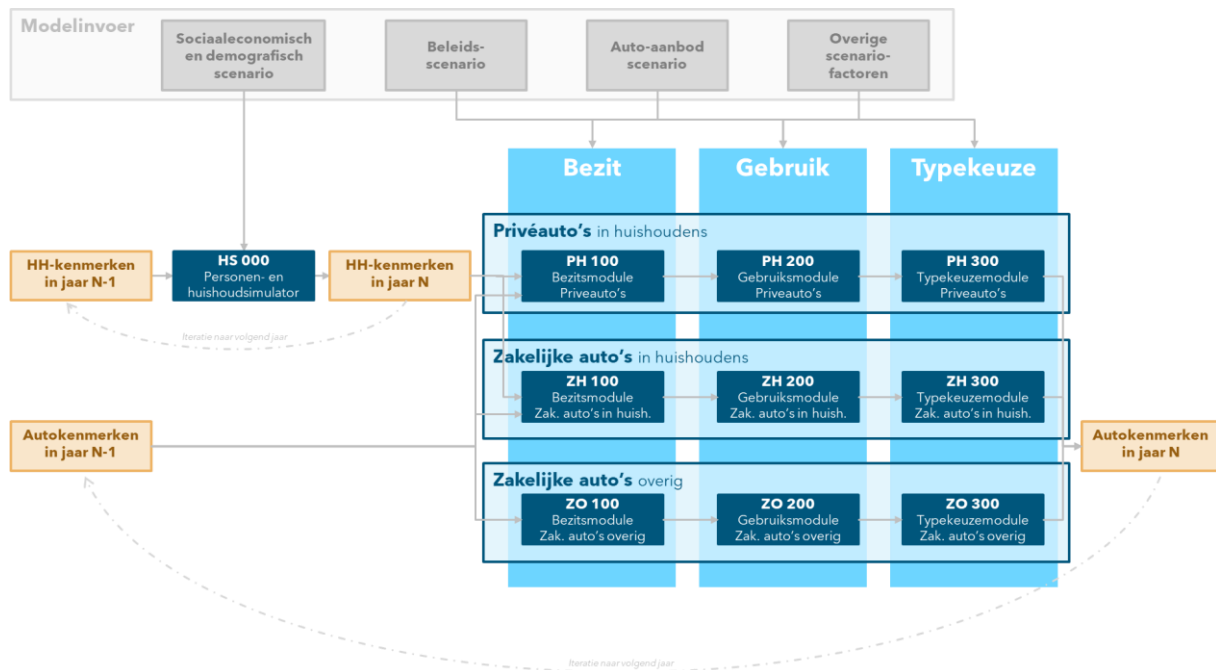
- Privéauto’s in huishoudens;
- Zakelijke auto’s in huishoudens;
- Overige zakelijke auto’s.

Binnen SPARK wordt het type van een auto bepaald in vijf dimensies:

- De groep waartoe het merk van de auto behoort (vijf categorieën: Duitse merken, Franse merken, Japans/Zuid-Koreaanse merken, overige merken en topmerken);
- Het marktsegment waarin de auto valt, hetgeen een maat is voor de grootte van de auto (vijf categorieën: A-segment, B-segment, C-segment, D-segment, E/F-segment);
- De leeftijd van de auto naar zes leeftijdsklassen: nieuw, 1-2 jaar oud, 3-5 jaar oud, 6-10 jaar oud, 11-15 jaar oud, 16 jaar en ouder);<sup>1</sup>
- De energiebron van de auto met zes categorieën: benzine, diesel, LPG, plugin hybride elektrisch (PHEV) met ook benzinemotor, PHEV-diesel en BEV. Voor scenario’s is FCEV als een zevende categorie beschikbaar (i.e. Fuel Cell Electric Vehicle, zoals bijvoorbeeld een waterstofauto);
- De importstatus van een auto met twee categorieën (wel/geen importauto).

Dat betekent dus dat er in de keuzemodellen 1800 autotypes onderscheiden kunnen worden binnen SPARK, hoewel deze niet altijd beschikbaar hoeven te zijn. De penetratie van PHEV en BEV wordt niet alleen gemodelleerd met discrete keuzemodellen voor de keuze op een bepaald moment, maar ook met diffusiemodellen voor de diffusie van nieuwe technologieën in de tijd. Daarnaast is de invloed van kenmerken van elektrische auto’s (niet alleen kosten maar ook actieradius en laadtijd) ook meegenomen via nieuw stated preference onderzoek (hypothetische keuze-experimenten) onder huishoudens.

<sup>1</sup> Dit betreft het typekeuzemodel (met 1800 keuze-alternatieven). In de simulaties van SPARK wordt wel een administratie bijgehouden waarbij altijd de precieze leeftijd van een auto bekend is op basis van het bouwjaar.



Toelichting modulecodes:

- HS: huishoudsimulator
- PH: privéauto's in huishoudens
- ZH: zakelijke auto's in huishoudens
- ZO: overige zakelijke auto's

Figuur 1. Structuur SPARK op hoofdlijnen

## Exogene invloedsfactoren, beleidsknoppen en uitvoer

SPARK modelleert de invloed van belangrijke invoervariabelen, zoals sociaal-demografische ontwikkelingen, bevolkingsspreiding en inkomensontwikkeling, EU bronbeleid, ontwikkeling van brandstofprijzen, elektriciteitsprijs, actieradius en laadinfrastructuur op autobezit, autogebruik en autotypekeuze. Speciale aandacht is er in het model voor de penetratie van volledig elektrische en plugin hybride elektrische (PHEV) voertuigen, maar bijvoorbeeld ook voor private lease en het belang van import en export.

Daarnaast kunnen met SPARK de effecten van (fiscale) beleidsmaatregelen in kaart gebracht worden. Hierbij gaat het om de volgende beleidsknoppen:

- mrb (inclusief provinciale opcenten);
- bpm;
- accijns;
- energiebelasting;
- btw;
- bijtelling;
- aanschafsubsidie voor nieuwe en/of tweedehands elektrische auto's;
- laadsubsidie;
- kilometerheffing;



- sloopsubsidie.

De uitvoer van SPARK betreft:

- Ramingen over de toekomstige omvang, samenstelling, gebruik en kosten (vaste en variabele) van het autopark;
- Effecten van beleidsmaatregelen hierop en een doorvertaling naar effecten op CO<sub>2</sub> praktijkuitstoot en emissies van luchtverontreinigende stoffen;
- Budgettaire effecten voor de overheid van beleidsmaatregelen.

### Werkwijze bij de schatting

Bij de schatting van deze modellen zijn wetenschappelijke standaarden gehanteerd om de kwaliteit zoveel mogelijk te garanderen. De gekozen aanpak voldoet aan de volgende kenmerken:

- De modellen worden zoveel mogelijk geschat op data over het in de werkelijkheid waargenomen gedrag van consumenten (empirische basis). Dit in tegenstelling tot het werken met veronderstelde coëfficiënten of het opleggen dat het goedkoopste alternatief in totale kosten van autobezit altijd gekozen wordt.
- Voor zover mogelijk zijn de modelschattingen uitgevoerd op registerdata (uit persoonsregister, huishoudregisters, baanregisters, de registerdata over auto's van de RDW) die beschikbaar en met elkaar te koppelen zijn bij het CBS. De schattingen vinden dan meestal plaats binnen de CBS-omgeving (CBS-server), of door export van niet-identificeerbare data naar buiten deze omgeving. Deze data zijn doorgaans van betere kwaliteit dan interviewdata (minder datafouten), hoewel in een enkel geval (b.v. detail van merk en model bij importauto's) bekend is dat er toch ook dataproblemen zijn. Een ander voordeel van gebruik van deze data is dat het niet over steekproeven gaat, maar over de populatie als geheel. Steekproeffouten worden zo vermeden en de datasets waarop geschat wordt zijn heel omvangrijk, zodat de nauwkeurigheid van de geschatte coëfficiënten vaak heel groot wordt (hoge t-waarden). Slechts waar deze CBS-data niet te gebruiken waren (b.v. gegevens over zakelijke auto's kunnen niet volledig aan de personen worden gekoppeld, of er zijn nog relatief weinig elektrische privéauto's) zijn additionele databronnen gebruikt (steekproeven en ook stated preference data).
- Gebruik van standaardsoftware met bewezen kwaliteit; SPARK is geprogrammeerd in de Python-software.
- Inzet van zeer ervaren mensen die de schattingen uitvoeren. Het schatten van modellen is vakwerk omdat bij elke stap keuzes moeten worden gemaakt in welke richting een model verder kan worden verbeterd. Ervaring over welke stappen het meest kansrijk zijn is daarbij een groot voordeel.
- Hanteren van wetenschappelijke statistische toetsen (met name de Likelihood-Ratio test) of een bepaalde modelaanpassing wel of geen modelverbetering is.
- Systematische aanpak bij specificatie-onderzoek. Aan het begin van een specificatieonderzoek zijn er zeer veel richtingen waarin je het model kunt ontwikkelen (toevoegen van interactievariabelen, toevoegen van niet-lineaire effecten, toevoegen van nest-coëfficiënten, etc.).
- Gebruik van logboek: alle modelschattingen worden bewaard. In een logboek is bijgehouden welke modellen zijn geschat en hoe deze beoordeeld zijn. Dit vormt de basis van de schattingsrapportage.
- Vier-ogenprincipe binnen het uitvoerende team: de schattingscode van elk type model is door iemand anders gecontroleerd op potentiële fouten. De ontwikkeling van elk type model is door iemand anders binnen het uitvoerende team gecontroleerd op basis van het logboek.

## 2.4 Uitkomsten voor de kwaliteit van de modelschatting

Een gedetailleerde bespreking van de schattingsuitkomsten is te vinden in de opgeleverde schattingsrapportage (Powerpoint-bestand) en alle gedetailleerde uitkomsten staan in het eveneens geleverde Excel-bestand.

Bij de modellen voor zakelijke auto's in huishoudens vinden we het volgende:

- Sterke modellen (goede fit) met veel (en vaak zeer) significante coëfficiënten;
- Tekens en grootte van de coëfficiënten zijn plausibel;
- t-ratio's (in absolute waarde) variëren tussen 2 (het minimum voor significantie bij 95% betrouwbaarheid) en 593. De hoge t-waarden zijn mogelijk door het schatten op honderdduizenden waarnemingen.

Bij de modellen voor overige zakelijke auto's vinden we het volgende:

- Sterke modellen (goede fit) met veel (en vaak zeer) significante coëfficiënten;
- Tekens en grootte van de coëfficiënten zijn plausibel;
- t-ratio's (in absolute waarde) variëren tussen 2 (het minimum voor significantie bij 95% betrouwbaarheid) en 706.

Bij de modellen voor privéauto's in huishoudens vinden we het volgende:

- Sterke modellen (goede fit) met veel (en vaak zeer) significante coëfficiënten;
- Tekens en grootte van de coëfficiënten zijn plausibel;
- t-ratio's (in absolute waarde) variëren tussen 2 (het minimum voor significantie bij 95% betrouwbaarheid) en 3000 (de hoogste waarde die wij ooit zijn tegengekomen). De hoge t-waarden zijn mogelijk door het schatten op miljoenen waarnemingen.

Zowel de experts namens de opdrachtnemer (zie in H1) als de opdrachtgever (zie eveneens in H1) waren zeer positief over de uitgevoerde modelschattingen en schattingsuitkomsten. Zo schreef prof. Dr. Harry Timmermans (naast twee punten met detailopmerkingen):

“Ik heb met veel genoegen de powerpoint en de bijbehorende excel-bestanden (met de schattingsresultaten – Significance) gelezen. Analyses zijn professioneel, vaak zijn varianten bekeken en vanuit de meeste geschatte parameters zijn conform verwachtingen. De meeste niet-significante parameters zijn ook interpreteerbaar. De goodness-of-fit van de modellen is overwegend redelijk tot goed.”

Prof. Dr. Henk Meurs schreef over de modelschattingen (naast opmerkingen over met name de toen nog uit te voeren validatie van de modelprognoses):

“De schattingen en beschrijving van de modelopzet en uitkomsten zien er i.h.a. goed uit en zijn goed te volgen. Coëfficiënten zien er qua teken en orde grootte plausibel uit.”

# 3. Kwaliteit van de modelimplementatie

## 3.1 Waar hangt de kwaliteit van de modelimplementatie van af?

De kwaliteit van de modelimplementatie hangt af van:

- Uitgebreide en heldere modelspecificatie;
- Systematische en transparante wijze van programmeren in de modelimplementatie, volgens van tevoren vastgelegde richtlijnen;
- Systematische softwaretesten.

## 3.2 Criteria voor de kwaliteit van de modelimplementatie

Het gaat bij de softwaretesten om de volgende testen:

### Automatische testen

De automatische tests, die zijn uitgevoerd, zijn te onderscheiden in unit-tests en integratie-tests.

Binnen het ontwikkelproces wordt met het principe gewerkt dat een ontwikkelaar voorafgaande aan het schrijven van de code voor een taak (unit-)tests definieert en opstelt, of dit tenminste doet gedurende het schrijven van de code. De ontwikkelaar start pas met de implementatie van een nieuwe taak wanneer de vorige succesvol, met geslaagde tests, is afgerond.

#### Unit-tests

Unit-tests dragen zorg voor de controle van de werking van een geïsoleerd stuk code, zoals een module, deelmodule of rekenstap. Zo kan gecontroleerd worden of een berekening met specifieke parameters een verwacht resultaat oplevert, of dat deze juist een verwachte, bedoelde, foutmelding of waarschuwing veroorzaakt. Door dit voor kleine code-eenheden, die 'units' worden genoemd, te testen ontstaat er vertrouwen in dit deel van de code. Ook na oplevering heeft dit voordelen, omdat de test faalt zodra een wijziging in de code onverwachte gevolgen heeft. Vervolgens kan dan redelijk nauwkeurig worden nagegaan waar eventuele onrechtmatigheden in de code zich bevinden.

De tests met betrekking tot de invoer (het inlezen van parameters, invoer en tussenresultaten) betreffen met name de vraag of het systeem het laat merken wanneer:

- Gegevens ontbreken;
- Aantallen niet overeenkomen met verwachte aantallen;
- Waarden niet consistent of incorrect zijn (zoals in het verkeerde format staan of buiten een gegeven bereik komen).

Incorrecte invoer moet tot correcte foutmeldingen of tenminste tot waarschuwingen leiden.

De tests met betrekking tot processen behelzen met name het controleren of uitkomsten van berekeningen voldoen aan de verwachtingen. Dit kunnen vooraf bepaalde exacte te verwachten resultaten zijn. Het kunnen ook sommaties zijn van resultaten zijn waarbij de delen moeten kloppen met het geheel. Tenslotte zijn er tests betreffende de richting waarin resultaten zouden moeten veranderen bij wijziging van invoer of parameters.

Bij zeer kleine te testen eenheden kan doorgaans eenvoudig bepaald worden wat de uitkomsten bij specifieke invoer zouden moeten zijn. Bij grotere eenheden, zoals hele modules, zijn soms resultaten beschikbaar uit de specificatiefase, waar prototypes zijn ontwikkeld en gebruikt. Daarnaast kan het voorkomen dat uit eerder uitgevoerde schattingsberekeningen invoerwaarden en verwachte uitvoerwaarden kunnen worden overgenomen.

Het is bij een applicatie van zekere omvang haast onmogelijk alle mogelijke scenario's te testen. Wel zijn alle regels aan code met unit-tests afgedekt.

### *Integratie-tests*

Integratie-tests zorgen ervoor dat de werking tussen verschillende onderdelen van een applicatie gecontroleerd wordt. Dit kunnen integratie-tests op module- en op applicatie niveau zijn.

De integratie-tests op moduleniveau controleren of het raamwerk een module kan aanroepen, de module correcte uitkomsten wegschrijft, de verwachte foutmeldingen op een correcte manier worden opgepikt en de volgende module de uitvoer van de vorige module correct oppikt.

De integratie-tests op applicatieniveau controleren of de gehele applicatie op basis van gegeven invoer, van begin tot eind, correct draait en de juiste uitvoer oplevert. Omdat deze tests lang kunnen duren is het streven om, qua omvang, beperkte test-datasets te hanteren. Voor deze datasets, die wel zoveel mogelijk situaties afdekken, worden, waar mogelijk, onafhankelijk van de applicatie (middels bijvoorbeeld Excel), de correcte uitvoerresultaten bepaald. Mocht het niet haalbaar zijn om de exacte uitvoer, onafhankelijk van de ontwikkelde applicatie, op applicatieniveau te bepalen, dan zijn er altijd nog de integratie-tests die controleren of bij veranderde invoer de uitvoer in een te verwachten plausibele richting wijzigt.

### **Handmatige testen**

De handmatige tests zijn de tests die door een persoon uitgevoerd moeten worden om te kijken of SPARK voor een gebruiker bruikbaar is. Deze tests laten zich niet goed vastleggen in een geautomatiseerd proces. Dit gaat bijvoorbeeld om het gebruik van de installer, en het uitvoeren van SPARK met behulp van de GUI.

### *Functionele test*

Nadat het model is geschat en de implementatie van de software is afgerond en met succes de andere tests heeft doorlopen, is het model ook functioneel getest. Hierbij zijn diverse scenario's met het model en gevoeligheidstests uitgevoerd. Doel van de functionele tests is om aan te tonen dat het model inhoudelijk plausibele resultaten genereert. De functionele test toont tevens aan dat aan de eisen met betrekking tot rekentijd, overdraagbaarheid en uitlegbaarheid van resultaten uit de vraagspecificatie voldaan wordt. De functionele testen maken deel uit van de validatiefase / proeftuin (zie H4), maar in de praktijk is het vaak nodig om te itereren tussen validatie en implementatie en de unit-, integratie- en systeem-testen van de software: validatie kan leiden tot aanpassingen in de software, die weer moeten worden getest en waarna de validatie opnieuw wordt uitgevoerd.

## **3.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelimplementatie**

De modelspecificatie die de ontwikkelaars/programmeurs gebruikten als basis voor hun werk is geschreven door degene die ook een belangrijke rol gespeeld heeft bij het opstellen van het conceptuele model, een groot deel van de modelschattingen zelf heeft uitgevoerd en de databewerkingen en andere modelschattingen heeft geleid. Deze modelspecificatie is zeer uitgebreid, zodat de ontwikkelaars zo precies mogelijk kunnen zien wat de bedoeling is. Ook vormt deze modelspecificatie de basis voor de technische documentatie van het model (het laatste is een product voor de opdrachtgever).

Teneinde de kwaliteit van de applicatie beheersbaar te maken, de overdraagbaarheid te vergroten en de kans op fouten te verkleinen, is een aantal richtlijnen gevolgd.

### *Benaming*

De namen van variabelen en klassen dienen helder en eenduidig te zijn en zelfs zonder documentatie al te verduidelijken wat de code doet. Dit betekent niet dat het hele zinnen moeten zijn, maar zeker ook niet een enkele letter.

### *Codedocumentatie*

Iedere klasse en iedere methode is voorzien te zijn van documentatie. Deze documentatie is met een specifiek format toegevoegd, zodanig dat met standaard instrumenten automatisch de systeemdokumentatie kon worden opgesteld (deze documentatie bestaat uit een korte afzonderlijk opgestelde inleiding, die de opzet en structuur beschrijft, en een groot automatisch gegenereerd deel).

Waar mogelijk is in de codedocumentatie met eenduidige namen en codes verwezen naar de in het ontwerp gedefinieerde (deel-)modules en dataverzamelingen.

#### *Lengte*

De lengtes van methodes en klassen zijn, voor wat betreft het aantal regels, beperkt. Methodes kunnen in een oogopslag worden overzien, dus passen op het scherm, teneinde ook andere ontwikkelaars snel inzicht te geven. Lengtes van de coderegels zijn ook beperkt tot bijvoorbeeld 80 karakters. Zie ook de richtlijnen die op <https://www.python.org/dev/peps/pep-0008/> worden genoemd.

#### *Beperkte complexiteit*

De complexiteit van een methode blijft binnen de perken. Eindeloos geneste if- en for- statements worden, om het overzicht te bewaren, vermeden. Beter is het om methoden op te splitsen in kleinere functionele eenheden.

#### *Geen herhaling*

Met het oog op het, in het geval van een bug, voorkomen dat op meerdere plekken dezelfde correctie moet worden doorgevoerd, wordt de herhaling van lappen code geminimaliseerd. Het werken met onder meer basisklassen, algemeen beschikbare methoden en extensie-methoden helpen bij het vermijden van herhaling van code.

#### *Tegenlezen*

Opgeleverde code is, zodra wijzigingen worden samengevoegd, door andere ontwikkelaars tegen gelezen. Ontwikkelde tests zijn keer op keer, na iedere wijziging en bij iedere samenvoeging met de rest van de code, uitgevoerd.

### 3.4 Uitkomsten voor de kwaliteit van de modelimplementatie

De werking van het model staat in detail beschreven in de geleverde technische documentatie. De uitgevoerde softwaretests en de uitkomsten ervan staan in het testrapport op GitLab. Hier is te zien dat (in een enkel geval na een of meer correcties) alle unit-tests en integratie-tests tot een positief resultaat hebben geleid.

SPARK als prognosemodel voor jaar-op-jaar prognoses voor bezit, gebruik en typekeuze van personenauto's wordt toegepast op een steekproef van ongeveer 200.000 individuele huishoudens. Via een expansieprocedure worden uitkomsten voor heel Nederland verkregen. SPARK is een echt microsimulatiemodel in die zin dat het de individuele huishoudens en individuele auto's van jaar op jaar volgt.

Ieder model is een abstractie van de complexe werkelijkheid, zo ook hier. Binnen SPARK geldt een aantal regels in de toepassing voor het maken van prognoses om de complexiteit en rekentijd binnen de perken te houden. Dit zijn vereenvoudigingen, die uitzonderingen en klein groepen en kleine effecten betreffen, terwijl de hoofdeffecten uit de werkelijkheid wel zijn opgenomen.

Voor personen/huishoudens geldt:

- Het aantal personen in een huishouden is maximaal 6;
- Het aantal personenauto's per huishouden (privé + zakelijk samen) is maximaal 6;
- Het aantal zakelijke personenauto's in een huishouden heeft wel effect op het aantal privéauto's maar niet andersom.

Binnen SPARK geldt voor auto's:

- Het automarktsegment van een auto is altijd A, B, C, D, of E. Voor de indeling worden de regels van RevNext gebruikt. Auto's die binnen deze regels worden toegewezen aan klasse "O" (i.e. overige, zoals bijv. campers) worden niet meegenomen omdat het in SPARK om personenauto's gaat;
- Als de energiebron alcohol is gebruiken we benzine als energiebron in SPARK;
- Zakelijke auto's in huishoudens zijn altijd leaseauto's;

- Zakelijke auto's in huishoudens komen altijd als nieuwe auto in een huishouden;
- Zakelijke auto's in huishoudens zijn nooit geïmporteerd;
- Een geïmporteerde auto is altijd minstens 1 jaar oud;
- Leaseauto's waarmee geen privékilometers gereden worden, of waarmee minder dan 500 privékilometers gereden worden, vallen in de categorie "zakelijke auto's overig" en niet in de categorie "zakelijke auto's in huishoudens", omdat ze niet onder de bijtellingsregeling vallen.

Binnen SPARK geldt voor de autohandel:

- Er wordt geen bedrijfsvoorraad gemodelleerd.



## 4. Kwaliteit van de modelvalidatie

Als onderdeel van de functionele test/validatie is een groot aantal testen met het geïmplementeerde model uitgevoerd. We noemen dit de ‘proeftuin’.

### 4.1 Waar hangt de kwaliteit van de modelvalidatie van af?

Nadat het model is geschat en in software geïmplementeerd en met succes de software tests heeft doorlopen is het model in de proeftuin functioneel getest/gevalideerd: werkt de software zoals deze behoort te werken, zijn er in de implementatie geen fouten gemaakt; zijn de uitkomsten van het model als geheel én op onderdelen conform de verwachtingen?

Hierbij zijn diverse scenario's met het model doorgerekend en er zijn gevoeligheidstests uitgevoerd. Het doel van de functionele tests is om aan te tonen dat het model inhoudelijk plausibele resultaten oplevert. Bij het bepalen van wat inhoudelijk plausibel is maken we gebruik van het in het begin van het project opgestelde toetsingskader (Memo 02, bijvoorbeeld ranges van elasticiteiten voor veranderingen in vaste en variabele kosten en inkomen). De functionele test is tevens bedoeld om te onderzoeken of voldaan wordt aan de eisen op het gebied van rekentijd, overdraagbaarheid en uitlegbaarheid van resultaten.

De proeftuinfase zelf ging van start met het opstellen van een plan van aanpak voor de validatie, dat is voorgelegd aan de opdrachtgever. Hierbij zijn ook de benodigde data besproken die in de validatie gebruikt worden om de uitkomsten van SPARK mee te vergelijken.

### 4.2 Criteria voor de kwaliteit van de modelvalidatie

Het gaat dan om deze testen:

#### 1. Testen voor de SPARK beginpopulatie voor eind 2018

Deze testen betreffen de vraag of het landelijke autopark, onderscheiden naar diverse deelmarkten en technische kenmerken op 31-12-2018, door de beginpopulatie van het model goed wordt weergegeven, d.w.z. overeenkomstig de waarnemingsgegevens voor dit moment (van het CBS, waaronder de RDW-data):

- Het landelijke autopark, onderscheiden naar deelmarkt (particuliere markt, zakelijke markt en private lease) en naar technische kenmerken (brandstofsoorten: benzine, diesel, LPG, PHEV-benzine, PHEV-diesel, BEV; autosegmenten en bouwjaren);
- Dit geldt niet alleen voor de aantallen auto's maar ook voor de autokilometers (jaarkilometrages) en voor de doorvertaling naar CO<sub>2</sub>-uitstoot en emissies van luchtverontreinigende stoffen (NO<sub>x</sub>, PM<sub>10</sub> en NH<sub>3</sub>);
- Verdeling van auto's en autokilometers over huishoudens.

#### 2. Testen voor de modeluitvoer voor de toekomst:

- Hierbij gaat het zowel om de modeluitkomsten voor zowel de komende jaren als voor jaren als 2030, 2040, 2050. Dit is uitgevoerd door het draaien van het model voor een bestaand lange termijn scenario, te kiezen in overleg met de opdrachtgever. De modeluitkomsten zijn dan vergeleken met prognoses van andere modellen voor hetzelfde scenario. Onderzocht is op deze manier of het model voldoende stabiel is op korte en lange termijn, maar ook in staat is om ingrijpende lange termijn veranderingen (zoals ingroei in het park van elektrische auto's) te simuleren. Om de overgangsfase vanuit de huidige modellen te ondersteunen zijn de doorrekeningen vergeleken met eerdere doorrekeningen op basis van Dynamo.
- Deze testen zijn niet alleen van belang voor het beoordelen van de plausibiliteit van de uitkomsten, maar ook voor het testen van het gebruiksgemak van het model en de runtijd. De benodigde inspanning voor het instellen van invoerparameters en runtijd van het model moeten zodanig beperkt zijn dat in korte tijd diverse scenario's doorgerekend kunnen worden.

Het is van groot belang dat model overdraagbaar is aan de opdrachtgever en de opdrachtgever het model zelfstandig kan toepassen.

**3. Testen van de gevoeligheid van het model op korte en lange termijn voor beleidsmaatregelen:**

Naast een goede weergave van het basisjaar en het referentiep pad voor de toekomst, zonder nieuw beleid, zijn in het model de effecten van (fiscale) beleidsmaatregelen op dezelfde indicatoren als hierboven bij onderdeel 1 beschreven geanalyseerd. De volgende potentiële veranderingen (beleidsknoppen) zijn hier onderzocht:

- ban op nieuwverkoop fossiele auto's vanaf 2035
- mrb
- bpm
- accijns
- aanschafsubsidie voor nieuwe en/of tweedehandsauto's
- kilometerheffing.

De uitkomsten van deze modelruns worden uitgedrukt in termen van absolute en relatieve effecten op het aantal auto's (per categorie) en autokilometers t.o.v. een referentiesituatie zonder deze maatregel. Ook is in de validatie gekeken naar uitkomsten van SPARK in de vorm van elasticiteiten, die zijn vergeleken met de range in het toetsingskader. Meer in het algemeen gesproken zijn de beleidseffecten uit het model beoordeeld op hun uitlegbaarheid.

**4.3 Uitgevoerde maatregelen en uitkomsten voor de kwaliteit van de modelvalidatie**

In het validatierapport zijn de uitkomsten gerapporteerd van de validatie van SPARK. Prognoses van SPARK zijn vergeleken met waarnemingsgegevens (bewerkte RDW-data) voor jaren die al achter ons liggen, met uitkomsten van het model Dynamo voor toekomstjaren en met de internationale literatuur voor elasticiteiten. Hierbij is het volgende naar voren gekomen.

De run tijd van SPARK voor een run tot en met 2040 is ongeveer 3 uur en tot en met 2060 ongeveer 9 uur.

De beginsituatie voor SPARK is die op 31-12-2018. Voor deze situatie is een bestand van huishoudens en gekoppelde auto's gecreëerd (via iterative proportional fitting IPF). Bij vergelijking van de uitkomsten hiervan met waarnemingen voor eind 2018 voor aantallen huishoudens en auto's naar belangrijke dimensies lijkt deze methode goed te functioneren.

Vervolgens is het model gekalibreerd op veranderingen in 2019. De uitkomsten hiervan geven aan dat ook dit werkt naar behoren.

Voor de jaren 2020 en 2021 is SPARK gedetailleerd bijgestuurd op de beschikbare waarnemingsgegevens voor deze jaren. De meest in het oog springende bijsturing is dat in 2020 en 2021 het autogebruik fors daalde door Covid-19 en lockdowns, wat natuurlijk niet in de prognoses van SPARK voor bijsturing zat, maar wat na bijsturing wel goed wordt voorspeld.

De groei van het totaal aantal auto's tot en met 2030 volgens SPARK komt iets lager uit dan de prognose van Dynamo bij hetzelfde WLO Hoog scenario. SPARK voorspelt een snellere penetratie van BEV's in het park dan Dynamo. In Dynamo is er in deze periode ook nog een snelle stijging van PHEV. SPARK levert voor 2030 een kleine toename van het gemiddelde kilometrage per auto, waar dit in Dynamo stabiel blijft.

SPARK en Dynamo zijn ook vergeleken voor 2040 en 2050, voor hetzelfde WLO Hoog scenario. Voor beide jaren is er een grote mate van overeenstemming van parkomvang en procentuele groei van het park vanaf 2018. Bij Dynamo weet de benzineauto zich langer te handhaven dan in SPARK, waar het marktaandeel van BEV snel groeit. Gerekend tot aan 2040 of 2050 is er een lichte stijging van het gemiddeld kilometrage per auto in SPARK.

De inkomenselasticiteit (alle effecten weerspiegeland of controlerend voor andere invloeden) in SPARK past in het bereik van de literatuur. Het effect van andere variabelen dan inkomen (zoals aantal huishoudens, banen, cohort-effect) op autobezit maakt ook deel uit van SPARK.

De kostenelasticiteiten van de parkomvang en het autogebruik zijn overgenomen uit andere Nederlandse modellen en bevinden zich binnen het bereik van de literatuur, zij het vooral aan de onderzijde van het bereik. Voor typekeuze zijn de effecten van de kostenvariabelen binnen SPARK op data van het CBS geschat en worden de resulterende elasticiteiten als plausibel gezien.

Er is een aantal beleidsruns uitgevoerd met SPARK. De uitkomsten hiervan zijn logisch/plausibel. Wel zijn de beleidseffecten in de diverse gevallen vrij gering. Met name de overgang naar BEV op de langere termijn vindt ook in de basisrun (zonder extra beleid) al plaats. Op de middellange termijn is er meer potentieel voor extra beleidsmaatregelen ter stimulering van BEV's.

# 5. Kwaliteit van de modelrapportage

## 5.1 Waar hangt de kwaliteit van de modelrapportage van af?

Bij aanvang van dit project is al besloten dat de modelrapportage transparant moet zijn: zo dienen zowel modelconcepten, de gebruikte data, de uitkomsten van de schattingen als de werking van het model als de uitkosten van software tests en validatie in detail te worden gerapporteerd aan de opdrachtgever (en voor een belangrijk deel ook aan de experts aan de zijde van opdrachtnemer en opdrachtgever). Dat dit leidt tot een aantal omvangrijke rapporten wordt op de koop toe genomen.

## 5.2 Criteria voor de kwaliteit van de modelrapportage

Het gaat bij kwaliteit van de modelrapportage om:

- Correctheid van de weergave;
- Volledigheid (in verband met de hierboven genoemde transparantie);
- Duidelijkheid, begrijpelijkheid en leesbaarheid (ook dit verhoogt de transparantie).

## 5.3 Uitgevoerde maatregelen voor een hoge kwaliteit van de modelrapportage

Alle conceptrapporten zijn standaard door iemand die niet auteur was beoordeeld op leesbaarheid en duidelijkheid. Deze opmerkingen worden meegenomen in de definitieve versie van het rapport. Alle rapporten zijn door de opdrachtgever gecontroleerd en diverse rapporten ook door de experts van beide zijden.

De geleverde rapporten betreffen:

Kwaliteitsrapportage	Dit rapport
Rapportage Datagebruik	Rapportages over verwerking en analyse van de aangeleverde gegevens en schattingsdata
Rapportage Modeltheorie	Rapportage conceptueel model
Rapportage Modelschatting	(1) Conceptrapportage specificatie onderzoek; (2) Rapportage van de schattingen: in deze rapportage worden zowel het zoekproces/ specificatie onderzoek, als ook de uiteindelijk geschatte modellen beschreven
Rapportage Modelvalidatie	Rapportage van de validatie- en testfase;
Rapportage Softwaretest	(1) Software documentatie, die direct uit de commentaren in de broncode kan gegenereerd worden (2) Samenvattende rapportage van alle software tests
Technische documentatie	Technische documentatie
Handleiding	Gebruiksvriendelijke handleiding van het personenautoparkmodel.

## 6. Conclusies

In de ontwikkeling van het instrument SPARK zijn er in de diverse modelstappen acties ondernomen om tot een kwalitatief goed product te komen. Het gaat dan per modelstap om de volgende activiteiten:

- **Modelspecificatie en -schatting:**
  - Een groot aantal inhoudelijke experts, toekomstige gebruikers en belanghebbenden is bij het proces betrokken;
  - Schatting van het model vond grotendeels plaats op bestanden met zeer veel waarnemingen (voor een belangrijk deel populatiebestanden) en op surveys toegesneden op de onderzoeksdoelstellingen van het project (waaronder stated preference onderzoek);
  - De schattingsresultaten zijn beoordeeld op plausibiliteit en statistische kwaliteitscriteria;
- **Modelimplementatie:**
  - De software die in dit project is geprogrammeerd is onderworpen aan een reeks van software-tests:
    - Unit-tests;
    - Integratie-tests;
    - Functionele tests;
- **Modelvalidatie:**
  - Het model is gekalibreerd en bijgestuurd op waarnemingsgegevens voor het autopark voor 2019, 2020 en 2021;
  - Het model is getest voor korte en lange termijn prognoses bij het WLO-Hoog scenario en vergeleken met prognoses van andere modellen en elasticiteiten uit de literatuur;
  - Diverse mogelijke beleidsmaatregelen zijn met het model doorgerekend en de uitkomsten zijn vergeleken met effecten uit andere modellen en elasticiteiten uit de literatuur;
- **Modelrapportage:**
  - Conceptversies van de diverse rapporten zijn tegen gelezen door reviewers bij zowel de uitvoerende als de begeleidende organisaties.



# significance

quantitative research

Grote Marktstraat 47  
2511 BH Den Haag  
Nederland

[info@significance.nl](mailto:info@significance.nl)  
+31 70 312 1530